



รายงานวิจัยฉบับสมบูรณ์

โครงการ แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ

โดย ดร. วีระศักดิ์ คุรุรัช และคณะ

เดือน กรกฎาคม ปี พ.ศ. ๒๕๔๒

รายงานวิจัยฉบับสมบูรณ์

โครงการ แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ

คณะผู้วิจัย

สังกัด

1. ดร. วีระศักดิ์ คุรุฑ์

มหาวิทยาลัยเทคโนโลยีมหานคร

2. รศ ดร. ใฮติยูกิ ทาคากิ

Kyushu Institute of Design

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

ชุดโครงการ แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ

สารบัญ

กิตติกรรมประกาศ	1
บทคัดย่อ.....	2
ABSTRACT.....	4
หน้าสรุปโครงการ (EXECUTIVE SUMMARY).....	6
เนื้อหางานวิจัย	8
OUTPUT	16
ภาคผนวก	17

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณสำนักงานกองทุนสนับสนุนการวิจัยที่ได้กรุณาให้ทุนสนับสนุนในการทำวิจัย
ชิ้นนี้ภายใต้สัญญาเลขที่ PDF/50/2540 นอกจากนี้มหาวิทยาลัยเทคโนโลยีมหานครยังได้อำนวย
ความสะดวกและสนับสนุนทุนในการเดินทางไปร่วมประชุมวิชาการต่างๆที่ต่างประเทศ และผู้วิจัย
ต้องขอขอบพระคุณท่าน รองศาสตราจารย์ ดร. ใฮติยุดิ ทาคากิ ที่ช่วยกรุณาส่งบทความที่จำเป็น
มาให้จากต่างประเทศพร้อมกับเสนอความคิดเห็นที่เป็นประโยชน์ต่องานวิจัยชิ้นนี้ ที่สำคัญอีกอัน
หนึ่งที่ทำให้งานนี้สำเร็จลุล่วงมาได้อีกระดับหนึ่งก็คือการเสียสละและการให้กำลังใจของครอบครัวที่
ทำให้ผู้วิจัยสามารถใช้เวลาที่มีค่าในการศึกษาค้นคว้าทำงานวิจัยชิ้นนี้

บทคัดย่อ

รหัสโครงการ: พีดีเอฟ/50/2540

ชื่อโครงการ: แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ

หน่วยงาน: ภาควิชาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีมหานคร

e-mail: werasak@mut.ac.th

ระยะเวลาโครงการ: 1 สิงหาคม 2540 – 31 กรกฎาคม 2542

วัตถุประสงค์: เสนอกรอบทฤษฎีที่ใช้เป็นแบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ ทั้งนี้ก็เพื่อเป็นแนวทางในการขยายขีดความสามารถของแบบจำลองข้อมูลเชิงสัมพันธ์ให้สามารถจัดการข้อมูลที่เกี่ยวข้องกับเวลาและข้อมูลที่คลุมเครือได้ ขีดความสามารถดังกล่าวจะช่วยทำให้การใช้งานของฐานข้อมูลมีขอบเขตกว้างมากขึ้นและนอกจากนี้ยังจะช่วยให้ระบบสารสนเทศมีความฉลาดมากขึ้นกว่าที่เป็นอยู่ในปัจจุบัน ที่จริงแล้วนักวิจัยหลายๆกลุ่มในสาขาระบบฐานข้อมูลได้เล็งเห็นความสำคัญของคุณลักษณะทั้งสองอย่างของข้อมูลมาเป็นเวลากว่าสิบปีแล้ว โดยคุณลักษณะทางด้านเวลาจะทำให้ระบบฐานข้อมูลที่มีมิติของเวลาสามารถจัดการประวัติข้อมูลหรือข้อมูลที่เปลี่ยนแปลงตามเวลาได้ ส่วนคุณลักษณะด้านความคลุมเครือจะทำให้ระบบฐานข้อมูลแบบคลุมเครือสามารถจัดการข้อมูลที่คลุมเครือซึ่งมีการใช้ใกล้เคียงกับชีวิตจริงได้

วิธีการ: หลักการที่ใช้ในงานวิจัยชิ้นนี้คือการนำเอาทฤษฎีของฟัซซีเซตมาใช้เป็นเครื่องมือในการจัดการความคลุมเครือของข้อมูล และนอกจากนี้ก็ได้มีการนำเอาแบบจำลองเวลาแบบไม่ต่อเนื่องมาใช้แนวความคิดของช่วงเวลาในแบบจำลองดังกล่าวได้ถูกนำแทนด้วยองค์ประกอบทั้งสามส่วนคือ เวลาเริ่ม เวลาสิ้นสุด และ ความยาวของช่วงเวลา ซึ่งองค์ประกอบแต่ละตัวจะถูกอธิบายด้วยเลขฟัซซีตัวหนึ่ง นอกจากนี้ ตัวสร้างข้อมูลเวลาสองชนิดในแบบจำลองนี้ได้ถูกศึกษา ชนิดหนึ่งเป็นแบบที่สร้างองค์ประกอบที่มีความหมายจากช่วงเวลาที่ถูกคลุมเครืออันหนึ่ง ส่วนอีกแบบหนึ่งจะสร้างช่วงเวลาคคลุมเครือหนึ่งช่วงจากข้อมูลที่ประกอบด้วยช่วงเวลาคคลุมเครือสองช่วง จากการศึกษาดังกล่าวพบว่า คุณสมบัติทางคณิตศาสตร์ของแบบจำลองเวลาที่คลุมเครือนี้เป็นไปตามที่ควรจะเป็น นอกจากนี้นงานวิจัยนี้ได้เสนอแนวความคิดเกี่ยวกับความสัมพันธ์ทางเวลาของเหตุการณ์ที่เกิดขึ้นด้วยเวลาที่ไม่ชัดเจนซึ่งจะเป็นสิ่งสำคัญสำหรับการสร้างและประมวลคำสั่งสอบถามฐานข้อมูล

หลังจากศึกษาและออกแบบแบบจำลองเวลาแบบคลุมเครือแล้ว โครงสร้างของแบบจำลองข้อมูลที่มีชื่อว่า FuzzTime ได้ถูกคิดค้นขึ้นมา โดย FuzzTime นี้เป็นแบบจำลองข้อมูลที่มีมิติของเวลาโดยสามารถจัดการข้อมูลแบบคลุมเครือได้ มันถูกสร้างขึ้นมาจากพื้นฐานของแบบจำลองข้อมูลเชิงสัมพันธ์แบบดั้งเดิม โครงสร้างของตารางใน FuzzTime ประกอบด้วยกลุ่ม attribute อยู่ 2 กลุ่มคือ กลุ่มของ attribute ทางเวลา และ กลุ่มของ attribute ที่ไม่เกี่ยวกับเวลา ในการออกแบบโครงสร้าง

และข้อบังคับบางอย่างนี้ได้มีการคำนึงถึงการรักษาแนวความคิดที่สำคัญของ key ซึ่งใช้อยู่ในแบบจำลองข้อมูลเชิงสัมพันธ์แบบดั้งเดิมด้วย นอกจากนี้งานชิ้นนี้ได้เสนอ operator ที่เกี่ยวข้องกับเวลาที่จะใช้ในการประเมินคำสั่งสอบถามในแบบจำลองนี้ด้วย

ผลการวิจัย: ทฤษฎีทางคณิตศาสตร์ของ แบบจำลองเวลาแบบคลุ่มเครือ และ แบบจำลองข้อมูลที่มีมิติเวลาและข้อมูลที่คลุ่มเครือ ซึ่งสิ่งใหม่ที่เกิดขึ้นคือความสามารถในการจัดการข้อมูลที่มีความคลุ่มเครือทั้งที่เกี่ยวกับเวลาและไม่เกี่ยวกับเวลาในกรอบทฤษฎีของแบบจำลองข้อมูลเดียวกัน

การถกและสรุป: แนวความคิดในเชิงทฤษฎีที่ถูกเสนอขึ้นมามีดังกล่าวนั้นเป็นแนวความคิดพื้นฐานของแบบจำลองที่สามารถจัดการข้อมูลที่มีทั้งความคลุ่มเครือและมิติของเวลา ผลลัพธ์ที่ได้เสนอออกมาสามารถที่จะใช้ในการพัฒนาต่อไปทั้งในทางทฤษฎีและในทางการพัฒนาขึ้นมาใช้งานจริง เช่น การศึกษาถึงทฤษฎีของคำสั่งสอบถาม ทฤษฎีการออกแบบฐานข้อมูลของแบบจำลองดังกล่าว เป็นต้น นอกจากนี้เทคนิคต่างๆที่ได้ถูกเสนอขึ้นมาในงานวิจัยชิ้นนี้ยังสามารถนำไปใช้ในด้านอื่นได้ด้วย เช่น ระบบสารสนเทศทางภูมิศาสตร์ (Geographic Information Systems) เป็นต้น แบบจำลองข้อมูลที่มีมิติของเวลานี้เป็นลักษณะที่เรียกกันโดยทั่วไปว่าเป็น attribute-time stamping งานวิจัยทางด้านนี้ยังมีสิ่งที่ต้องทำอีกมาก

ข้อแนะนำ: สิ่งที่จะต้องทำต่อไปสำหรับงานนี้คือการศึกษาลึกถึงทฤษฎีของคำสั่งสอบถาม เช่น การขยายขีดความสามารถของรีเลชันแนลอัลจิبرا และ การศึกษาถึงความซับซ้อนของภาษาและแบบจำลอง ทั้งนี้เพื่อนำไปสู่การพัฒนาต้นแบบในรูปแบบของซอฟต์แวร์ DBMS ได้

คำสำคัญที่เกี่ยวข้อง: แบบจำลองข้อมูลที่มีมิติเวลา, แบบจำลองข้อมูลแบบคลุ่มเครือ, เวลาที่คลุ่มเครือ, ข้อมูลเวลา, ข้อมูลคลุ่มเครือ

ABSTRACT

Project Code: PDF/50/2540

Project Name: A Temporal Database Model with Fuzzy Knowledge of Time

Organization: Department of Computer Engineering, Mahanakorn University of Technology

e-mail: werasak@mut.ac.th

Project Period: August 1, 1997 – July 31, 1999

Objective: This research has proposed a formal framework to incorporate fuzzy temporal knowledge into a temporal database model. The major purpose of this research is to extend the capability of the relational data model to be able to handle both temporal data and fuzzy data. Such capability will help an expansion of database applications, and also will increase the intelligence of information systems. The importance of the two features, time and fuzziness, of data has been widely recognized by researchers in the field of databases for nearly two decades. The time dimension allows temporal database systems to handle historical data or time-varying data inherently. On the other hand, the fuzziness aspect allows fuzzy databases to incorporate vague data.

Methodology: In our approach, the fuzzy set theory has been employed as a theoretical tool to handle the fuzziness feature of data, and the discrete time model has been assumed. The concept of a time interval in the traditional discrete time model has been extended by representing it with three components, the starting time, the ending time and the duration, each of which is modeled by a fuzzy number. Then, both unary and binary temporal constructors, which are applied to one and two fuzzy time intervals respectively to produce another interval, have been studied. It has been proven that mathematical properties of the fuzzy time model have been satisfied. In addition, fuzzy temporal relationships among events, which will be useful in specifying and evaluating queries, have been defined and their characteristics have been investigated.

Based on the relational data model and the fuzzy discrete time model above, we have proposed a structure of a fuzzy temporal data model called FuzzTime. A fuzzy temporal relation in FuzzTime consists of two main parts – a group of temporal attributes and a group of non-temporal attributes. The design of this structure and some constraints has been accomplished by an effort of conserving the importance concept of key in the traditional

data model. Moreover, time related operators for query evaluation of the model have been proposed.

Research Result: Theoretical frameworks of a fuzzy time model and a temporal data model with fuzzy knowledge of time have been proposed. These novel concepts and techniques in databases allow the ability to represent and to manage both fuzzy temporal data and fuzzy non-temporal data in a homogeneous framework.

Discussion and Conclusion: The proposed theoretical notion is to be used for managing fuzzy and temporal data. The result can be a basis for further investigation and development both in theory and in practice. For instance, a new query language and a theory of database design based on our model need to be introduced. In addition, novel techniques proposed in this work can be employed in other areas such as geographic information systems. In fact, there are plenty of rooms for further research in this direction.

Recommendation: The next step that can be carried out is to study a theory of query languages and their evaluations such as an extension of the traditional relational algebra and the study of complexity of the languages and the model. This will lead to an implementation of a DBMS software prototype.

Keywords: Temporal Data Model, Fuzzy Data Model, Fuzzy Time, Temporal Data, Fuzzy Data

หน้าสรุปโครงการ (EXECUTIVE SUMMARY)

1. ชื่อโครงการ "แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ"

"A Temporal Database Model with Fuzzy Knowledge of Time"

2. ชื่อหัวหน้าโครงการ หน่วยงานที่สังกัด ที่อยู่ หมายเลขโทรศัพท์ โทรสาร และ e-mail

ดร. วีระศักดิ์ คุรุฑ์

ภาควิชาวิศวกรรมคอมพิวเตอร์

มหาวิทยาลัยเทคโนโลยีมหานคร

51 ถนนเชื่อมสัมพันธ์ หนองจอก กรุงเทพฯ 10530

โทรศัพท์ 9883655, 9883666 ต่อ 228 โทรสาร 9883687

e-mail. werasak@mut.ac.th

3. สาขาวิชาที่ทำการศึกษาวิจัย

Information Sciences and Artificial Intelligence

4. งบประมาณ

งบประมาณทั้งโครงการ 355,953.53 บาท

5. ระยะเวลาดำเนินการ

2 ปี

6. ได้นำโครงการนี้ หรือโครงการที่มีส่วนเหมือนกับเรื่องนี้บางส่วน เพื่อขอทุนต่อแหล่งทุน อื่นที่ใดบ้าง

ไม่ได้เสนอ

7. ปัญหาที่ทำการวิจัย และความสำคัญของปัญหา

ในการใช้งานของระบบฐานข้อมูลโดยทั่วไปนั้น เราจะพบได้ว่า รูปแบบของเวลา (Time) และ ความคลุมเครือ (fuzziness) ของข้อเท็จจริง เป็นคุณลักษณะสำคัญที่มีอยู่ในข้อมูลที่จะเก็บไว้ในฐานข้อมูล เช่น ระบบฐานข้อมูลที่เกี่ยวข้องกับอาชญากรรม ระบบฐานข้อมูลที่เกี่ยวข้องกับประวัติผู้ป่วย และระบบฐานข้อมูลที่ใช้ในการวิเคราะห์ข้อมูลทางเศรษฐกิจ เป็นต้น แต่ในปัจจุบันนี้ระบบบริหารฐานข้อมูล (DBMS) จะอาศัยแบบจำลองข้อมูลเชิงสัมพันธ์ และ แบบจำลองข้อมูลแบบดั้งเดิมชนิดอื่นเป็นพื้นฐาน ซึ่งแบบจำลองเหล่านี้ไม่สามารถที่จะสนับสนุนให้ระบบฐานข้อมูลสามารถจัดการเก็บความรู้ที่เกี่ยวกับเวลาและความไม่สมบูรณ์ของข้อมูลได้ ปัญหานี้ทำให้การใช้งานของระบบฐานข้อมูลถูกจำกัดอยู่ในวงแคบ และไม่ได้เป็นการจำลองที่จะสะท้อนต่อสิ่งที่เกิดขึ้นจริงในระบบที่ถูกจำลองโดยฐานข้อมูลนั้นๆ

8. วัตถุประสงค์

- เพื่อเสนอ แบบจำลองข้อมูล(data model) ที่มีขีดความสามารถในการจัดการทั้ง ข้อมูลที่มีมิติของเวลา (temporal data) และ ข้อมูลที่คลุมเครือ (fuzzy data)
- เพื่อเสนอเทคนิคในเชิงคณิตศาสตร์ที่อาจจะนำไปใช้ในการพัฒนา intelligent information system ได้
- เพื่อแสดงให้เห็นถึงความสำคัญและปฏิสัมพันธ์ที่มีต่อกันของข้อมูลทั้ง 2 ชนิด

9. การดำเนินการวิจัย

1. ศึกษางานที่มีอยู่ (ส.ค. 40 - ธ.ค. 40)
2. วิเคราะห์คุณสมบัติของแบบจำลองที่มีอยู่ (ม.ค. 40 – เม.ย. 41)
3. กำหนดคุณลักษณะของแบบจำลองที่จะเสนอ (พ.ค. 41 - ส.ค. 41)
4. ออกแบบ model และศึกษาคุณสมบัติของมัน (ก.ย. 41 - ก.พ. 42)
5. ศึกษาถึงการจัดการการขาดหายไปของข้อมูลเวลาในแบบจำลอง (มี.ค. 42 - ก.ค. 42)

เนื้อหางานวิจัย

บทนำ

ระบบจัดการฐานข้อมูล (DBMS) ที่แพร่หลายในปัจจุบันได้ใช้แบบจำลองข้อมูลเชิงสัมพันธ์ [9] ซึ่งข้อดีของมันก็คือความง่ายของแบบจำลอง และการมีทฤษฎีพื้นฐานทางคณิตศาสตร์มารองรับ แต่อย่างไรก็ตามข้อจำกัดของแบบจำลองข้อมูลดังกล่าวก็มีเช่น มันไม่สามารถที่จะจัดการข้อมูลที่มีความไม่สมบูรณ์ได้ [11, 12] นอกจากนี้ DBMS ที่สนับสนุนแบบจำลองดังกล่าวในปัจจุบันไม่สามารถจัดการข้อมูลที่มีมิติของเวลาได้โดยสะดวกต้องอาศัย application program ในการทำงานดังกล่าว [10] ดังนั้นการวิจัยในการขยายแบบจำลองข้อมูลเชิงสัมพันธ์เพื่อรองรับขีดความสามารถดังกล่าว จึงได้รับความสนใจจากนักวิจัยทางด้านระบบฐานข้อมูลเป็นอย่างมาก อย่างไรก็ตามการวิจัยที่ผ่านมาแบ่งเป็น 2 ทิศทางที่เป็นอิสระต่อกัน คือ งานทางด้านฐานข้อมูลที่คลุมเครือ [11, 13] และ ทางด้านฐานข้อมูลที่มีมิติของเวลา [15, 16] มีงานของนักวิจัยบางกลุ่มที่ได้ศึกษาถึงเรื่องของเวลาและความคลุมเครือ แต่โดยส่วนใหญ่แล้วจะเป็นงานที่มีมุมมองมาจากทางด้าน Artificial Intelligence เช่น งานของ Dutta [6, 7] ที่เสนอ fuzzy temporal logic ซึ่งแสดงถึงความสัมพันธ์ทางเวลาของเหตุการณ์ที่คลุมเครือ และ งานของ Dubois และ Prade [4] ซึ่งได้เสนอกรอบทฤษฎีสำหรับการ process ความรู้ทางเวลาที่คลุมเครือ โดยเน้นไปที่การอ้างเหตุผลทางเวลา นอกจากนี้ก็มี Barro และคณะ [2] ที่เสนอแบบจำลองและภาษาในการจัดการเวลาที่คลุมเครือแต่เป็นการเน้นไปที่การใช้งานในระบบผู้เชี่ยวชาญ ทางด้านระบบฐานข้อมูลเองก็มีนักวิจัยไม่กี่กลุ่มที่ให้ความสนใจแก่ข้อมูลเวลาที่ไม่นั่นนอนในแบบจำลองข้อมูลที่มีมิติของเวลา เช่น Dyreson และ Snodgrass [8] ได้ศึกษาความไม่นั่นนอนของเวลาที่เกิดจากการสุ่ม (random) ในแบบจำลองข้อมูลที่เป็นแบบเชิงสัมพันธ์ และใช้ทฤษฎีความน่าจะเป็นในการแก้ปัญหา ส่วน De Tre และคณะ [3] ได้ศึกษาการจำลองข้อมูลเวลาที่คลุมเครือในระบบฐานข้อมูลเชิงวัตถุ โดยใช้แบบจำลองเวลาที่คลุมเครือ [17] เป็นพื้นฐาน

งานวิจัยชิ้นนี้เกิดจากการตระหนักถึงข้อเท็จจริงที่ว่าความคลุมเครือเกิดขึ้นเสมอในข้อมูลที่เป็นเวลา เช่น คุณสมชายเข้ามาทำงานเวลาประมาณ 8.30 น หรือ บ้านหลังนี้ถูกโจรกรรมเมื่อเวลาระหว่างประมาณ 3.00 น. ถึง ประมาณ 4.00 น. นอกจากนี้ ระบบสารสนเทศหลายๆประเภทจำเป็นที่จะต้องเก็บข้อมูลที่คลุมเครือและมีมิติเวลาที่คลุมเครือด้วย เช่น ระบบสารสนเทศทางอาชญากรรม (Criminal Information Systems) หรือ ระบบสารสนเทศทางการแพทย์ (Medical Information Systems) หรือ ระบบสารสนเทศทางภูมิศาสตร์ (Geographic Information Systems) ดังนั้นผลงาน

วิจัยที่เสนอขึ้นมานี้เป็นเทคนิคใหม่อันหนึ่งที่จะจัดการข้อมูลที่คลุมเครือและมีมิติของเวลาที่คลุมเครือในแบบจำลองข้อมูลเชิงสัมพันธ์ ซึ่งใช้เป็นพื้นฐานสำหรับระบบสารสนเทศต่างๆ

วิธีการทดลอง

งานวิจัยนี้เป็นงานวิจัยเชิงทฤษฎี ดังนั้นในส่วนนี้จะเป็นการกล่าวถึงแนวทางและลำดับในการเสนอผลงานที่จะกล่าวถึงในหัวข้อถัดไป

1. หลังจากการศึกษาถึงงานที่เกี่ยวข้องที่ผ่านมา เราต้องเสนอรูปแบบโครงสร้างของแบบจำลองเวลาแบบคลุมเครือ (fuzzy time model) และแบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ (fuzzy temporal data model) ทั้งนี้เพื่อแสดงให้เห็นถึงโครงสร้างที่เหมาะสมในการเก็บและจัดการข้อมูล บรรทัดฐานที่ใช้ในการเสนอมีดังนี้
 - 1.1 ต้องเป็นรูปแบบที่มีพื้นฐานมาจากแบบจำลองข้อมูลเชิงสัมพันธ์ ซึ่งมีทฤษฎีทางคณิตศาสตร์ที่แข็งแกร่งมารองรับ
 - 1.2 ต้องมีพื้นฐานมาจากแบบจำลองเวลาแบบไม่ต่อเนื่อง (discrete time model) ซึ่งเป็นที่ยอมรับโดยทั่วไปว่าเหมาะสมที่จะใช้กับทฤษฎีทางฐานข้อมูล
 - 1.3 ต้องเป็นลักษณะของการประทับเวลาในระดับทUPLE (tuple time stamping model) เนื่องจากวิธีการนี้จะทำให้แบบจำลองข้อมูลเป็นไปตามกฎของ normal form ที่ 1 และง่ายต่อการเข้าใจ
 - 1.4 ข้อมูลเวลาที่สนใจเป็นข้อมูลเวลาของเหตุการณ์ที่เกิดขึ้นจริงในโลกที่เราจำลอง (valid time) [14]
2. ศึกษาถึงข้อบังคับเพื่อให้สอดคล้องกับทฤษฎีดั้งเดิม ข้อบังคับที่สำคัญที่สุดคือเรื่องของ key นั่นก็คือว่า ทUPLE แต่ละตัวในรีเลชันเดียวกันต้องมีค่าของ key ที่ไม่เหมือนกัน
3. เสนอ operator ต่างๆที่จะใช้ในการจัดการข้อมูลเชิงเวลา ซึ่งก็มี operator ที่ใช้ในการสร้างข้อมูลเวลาอันหนึ่งจากข้อมูลเวลาตัวอื่น และ operator ที่ใช้ในการดึงข้อมูลจากฐานข้อมูลตามเงื่อนไขของเวลา
4. ศึกษาถึงความสัมพันธ์ทางเวลาตามทฤษฎีของ Allen [1] แต่ขยายแนวความคิดให้จัดการกับข้อมูลทางเวลาที่คลุมเครือได้ ความสัมพันธ์ทางเวลานี้จะถูกนำมาใช้ในการกำหนดเงื่อนไขทางเวลาของข้อเท็จจริงเพื่อดึงข้อมูลออกมา

ผลการทดลอง

จากแนวทางและขอบเขตที่วางไว้ในหัวข้อที่แล้ว ผลลัพธ์ที่ได้จากการวิจัยและออกแบบ สามารถสรุปออกมาได้ดังนี้

1. แบบจำลองเวลาแบบคลุมเครือ

แบบจำลองเวลาแบบไม่ต่อเนื่องนั้นจะแบ่งเส้นเวลา (time line) เป็นช่วงที่เล็กๆ โดยถือว่าเหตุการณ์ที่เกิดขึ้นในช่วงเล็กๆเดียวกันเป็นเหตุการณ์ที่เกิดขึ้นพร้อมกัน นอกจากนี้ได้มีการสมมุติว่า ช่วงเล็กๆทั้งหมดนั้นสมนัยกับเลขจำนวนเต็มเริ่มจาก 0, 1, 2, ... ไปเรื่อยๆ ดังนั้นช่วงเวลาเล็กๆทั้งหมดนั้นมีการเรียงลำดับแบบเชิงเส้น เราเรียกช่วงเล็กๆเหล่านั้นว่า โครนอน (chronon) [16] ในแบบจำลองเวลาแบบคลุมเครือนี้ เรากำหนดว่า ช่วงเวลาเป็นหน่วยเวลาพื้นฐาน โดยช่วงเวลาใดๆประกอบด้วยข้อมูลต่อไปนี้

1. จุดเริ่มต้นของช่วงเวลา ซึ่งเป็นพีชชีเซตของโครนอน หรืออีกนัยหนึ่งคือเป็นพีชชีเซตของเลขจำนวนเต็ม
2. จุดปลายของช่วงเวลา ซึ่งเป็นพีชชีเซตเช่นเดียวกันกับจุดเริ่มต้นของช่วงเวลา
3. ความยาวของช่วงเวลา ซึ่งกำหนดเป็นช่วงของความยาวที่เป็นไปได้

ตัวอย่างเช่น คุณสมชายเข้ามาทำงานประมาณ 8.00 น และเลิกประมาณ 17.00 น เป็นเวลาทั้งหมด 7 ถึง 9 ชั่วโมง เป็นต้น เราเรียกรูปแบบการเก็บข้อมูลเวลาดังด้วยองค์ประกอบทั้งสามว่าเป็น temporal specification และใช้สัญลักษณ์ดังนี้

$$T = \langle [s, e], c_1 \leq d \leq c_2 \rangle$$

โดยที่ s คือจุดเริ่มต้นของช่วงเวลา e คือจุดปลายของช่วงเวลา และ c_1 และ c_2 คือช่วงของความยาวที่สั้นที่สุดและยาวที่สุดตามลำดับ

จากโครงสร้างดังกล่าว เราสามารถสร้าง unary temporal constructor ได้ 3 ตัวดังนี้คือ

$$\begin{aligned} \text{end}(T) &= [(e - c_1 + 1) \cup (e - c_2 + 1)] \cap s \\ \text{start}(T) &= [(s + c_1 - 1) \cup (s + c_2 - 1)] \cap e \\ \text{dur}(T) &= (e - s + 1) \cup \{d_1 \mid d_1 \text{ satisfies } C(d)\} \end{aligned}$$

โดยที่ $\cup$ หมายถึง convex hull union operation [5] และ $\cap$ หมายถึง intersect operation เราจะเห็นได้ว่าทั้งสาม constructor นี้เป็นการสร้างข้อมูลเวลาที่มีความหมายชัดเจนขึ้นจากข้อเท็จจริงของเวลาที่มีอยู่ในรูปของ temporal specification

นอกจากนี้ งานนี้ได้มีการเสนอเทคนิคการประมวลผลของ binary temporal constructor ซึ่งสามารถสรุปได้ดังนี้

$$\begin{aligned}\mu_{\text{start}(T1) \cup (T2)}(n) &= \sup_x \mu_{\text{start}(T1) \cup (T2)}([n, x]) \\ \mu_{\text{end}(T1) \cup (T2)}(n) &= \sup_x \mu_{\text{end}(T1) \cup (T2)}([x, n]) \\ \mu_{\text{union}(T1) \cup (T2)}(d) &= \sup_{|d-n1-n2| \leq 1} \mu_{\text{union}(T1) \cup (T2)}([n1, n2])\end{aligned}$$

โดยที่ $\cup$ คือ binary temporal constructor ซึ่งเป็นการ union หรือ การ intersect ของ temporal specification 2 ตัว คุณสมบัติทางคณิตศาสตร์ของ $\cup$ คือ

- 1 Identity
- 2 Associative
- 3 Commutative
- 4 Distributive

สำหรับความสัมพันธ์ทางเวลาได้ถูกนำเสนอขึ้นมาโดยใช้ความสัมพันธ์ระหว่างช่วงเวลาที่ถูกเสนอขึ้นมาจากงานของ Allen ซึ่งมีทั้งหมด 13 ความสัมพันธ์ แต่ในกรณีของงานนี้ช่วงเวลาอยู่ในรูปของ temporal specification ข้างต้น ดังนั้นระดับของความสัมพันธ์จะถูกวัดออกมาเป็น degree เช่น

$$\begin{aligned}\text{Poss}(T1 \text{ precede } T2) &= \text{Poss}(\text{end}(T1) \text{ precede } \text{start}(T2)) \\ \text{Nec}(T1 \text{ precede } T2) &= \text{Nec}(\text{end}(T1) \text{ precede } \text{start}(T2))\end{aligned}$$

โดยที่ Poss และ Nec คือ possibility และ necessity function [5] ตามลำดับ ส่วนความสัมพันธ์อื่นๆเช่น during, equal, meet, overlap, start และ finish ก็สามารถประเมินได้เช่นเดียวกันกับข้างต้น

2. แบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือ (FuzzTime)

เราสามารถนำโครงสร้างของแบบจำลองเวลาแบบคลุมเครือที่เสนอออกมาข้างต้นมาดัดแปลงใช้กับแบบจำลองข้อมูลเชิงสัมพันธ์ที่มีมิติของเวลา เพื่อให้ได้แบบจำลองที่เราต้องการ ซึ่งจะเรียกว่า FuzzTime

relation schema ของแบบจำลองประกอบด้วย set ของ attribute ซึ่งโครงสร้างของแต่ละ attribute สามารถนิยามได้โดยอาศัยแนวความคิดด้านตัวแปรทางภาษา (linguistic variables) [18] ดังนี้

$$\text{Attr} = (A, U, T, G, M)$$

โดยที่ A คือชื่อของ attribute, U เป็น data domain พื้นฐาน, T เป็น domain ที่เป็น linguistic term, G เป็น syntactic rule และ M คือ semantic rule

$$M : U \cup T \rightarrow [0, 1]^U$$

โดยที่ $M(u) = \{1.0/u\}$ สำหรับทุกค่าที่ $u \in U$ ค่าใดๆที่เป็นสมาชิกของ $U \cup T$ สามารถถูกกำหนดให้เป็นค่าของ attribute A ได้ นั่นคือ

$$\text{dom}(A) = U \cup T$$

temporal relation instance r ของ schema R ซึ่งประกอบด้วย n attribute ($A_1, A_2, \dots, A_n$) สามารถนิยามได้ดังนี้

$$r \subseteq \times_{j=1}^n \text{dom}(A_j) \times_{k \in \{r, s, e, t, u\}} \text{dom}(A_k)$$

โดยที่ A_r คือ reincarnation attribute A_s คือ attribute ที่แสดงถึงเวลาเริ่มต้น A_e คือ attribute ที่แสดงถึงเวลาที่สิ้นสุด A_t คือ attribute ที่แสดงถึงช่วงเวลาต่ำสุดที่เป็นไปได้ และ A_u คือ attribute ที่แสดงถึงช่วงเวลาสูงสุดที่เป็นไปได้ เราจะเห็นได้ว่า attribute เหล่านี้ทั้งหมดเกี่ยวข้องกับเวลา

key ที่ไม่แปรเปลี่ยนตามเวลา (TIK) ซึ่งก็คือ key ของ snapshot relation ของ temporal relation ดังนั้นต้องมีกฎข้อบังคับว่าค่าของ attribute ทั้งหมดที่เป็น TIK ต้องเป็นค่าที่แน่นอน

นั่นคือ $T = \emptyset$ ดังนั้น temporal relation ทุกตัวจะมี key ที่สามารถใช้ค่าของมันในการชี้เฉพาะถึง tuple แต่ละ tuple ได้ แนวความคิดของ key สามารถอธิบายได้อีกนัยหนึ่งคือ

สมมุติว่า t_1 และ t_2 เป็น 2 ทัวเพิลใดๆใน temporal relation และ T_1 และ T_2 เป็น temporal specification ของมันตามลำดับ เราจะได้ว่า ถ้า $t_1[TIK] = t_2[TIK]$ และ $t_1[A_i] \neq t_2[A_i]$ แล้ว T_1 และ T_2 ต้องไม่มีความเป็นไปได้ของการ overlap กัน

โดยอาศัยทฤษฎี possibility เราสามารถที่จะแปลค่าที่คลุมเครือของ attribute ใดๆว่าเป็น possibility distribution โดยทั่วไป tuple แต่ละ tuple ใน relation instance จะมีความหมายในลักษณะเดียวกัน นั่นคือ เป็น possibility distribution ของ tuple ที่เป็นไปได้ นั่นคือ

$$M(t) = \{\mu(t_1)/t_1, \mu(t_2)/t_2, \dots, \mu(t_m)/t_m\}$$

โดยที่ t คือ tuple ใน relation instance, $\mu(t_i) = \inf_j \text{Poss}(t[A_j] = t_i[A_j])$, $t[A_j]$ และ $t_i[A_j]$ เป็นค่าของ attribute A_j ใน tuple t และ t_i ตามลำดับ

ในงานวิจัยชิ้นนี้ เราได้กำหนด operation สำหรับการดึงข้อมูลเวลาออกจากฐานข้อมูลโดยการกำหนดเงื่อนไขของเหตุการณ์ ดังนี้

$$(\varphi(\text{cond} : r) = \{ \langle ([A_s], [A_e]), t[A_i] \leq d \leq t[A_u] \rangle, \mu \mid t \in r, \mu = \min(\mu_{\text{cond}}(t), t[A_\mu]) \})$$

$$(\varphi(\text{cond} : r) = \{ \langle [\text{start}[T], \text{end}[T]], \text{dur}(T) \rangle, \mu \mid t \in r, \mu = \min(\mu_{\text{cond}}(t), t[A_\mu]) \})$$

นอกจากนี้ยังได้มีการพิสูจน์ว่า ผลลัพธ์จาก operation ข้างต้นเป็นไปตามกฎข้อบังคับของ key

บทวิจารณ์

ผลของงานวิจัยชิ้นนี้คือกรอบทฤษฎีของแบบจำลองข้อมูลที่มีมิติของเวลา ซึ่งที่เสนอออกมาคือโครงสร้างของแบบจำลอง รวมทั้งกฎข้อบังคับพื้นฐานที่เกี่ยวข้องกับ key และ operation ต่างๆที่ใช้สำหรับจัดการกับข้อมูล งานชิ้นนี้ล้ำกว่าแผนงานที่วางไว้ตั้งแต่เริ่มโครงการ ดังนั้นในส่วนของการศึกษาถึงองค์ประกอบที่จะมีผลต่อประสิทธิภาพต่อการสร้าง software ดันแบบจึงยังไม่ได้ศึกษา นอกจากนี้เนื้อที่จะต้องวิจัยในเชิงทฤษฎียังมีอีกมากก่อนที่จะ implement ขึ้นมาใช้งานได้ เช่น ต้อง

ออกแบบภาษาที่จะนำมาใช้เป็นทั้ง manipulation language และ data description language แต่อย่างไรก็ตาม เราสามารถสร้าง module ขึ้นมาไว้ระหว่าง RDBMS ที่ใช้กันอยู่ในปัจจุบันกับผู้ใช้ โดย module นี้จะทำหน้าที่แปลงแบบจำลองข้อมูลที่มีมิติของเวลาแบบคลุมเครือที่ผู้ใช้ให้อยู่ให้เป็นแบบจำลองข้อมูลเชิงสัมพันธ์ที่ถูก support โดย RDBMS ดังกล่าว

เอกสารอ้างอิง

1. Allen J. F., Maintaining knowledge about temporal intervals, CACM, Vol. 26, 1983, 832-843.
2. Barro S., Marin R., Mira J. and Paton A. R., A model and a language for the fuzzy representation and handling of time, Fuzzy Sets and Systems 61, 1994, 153-175.
3. De Tre G., De Caluwe R., Van Der Cruyssen B. and Van Gyseghem N., Towards Temporal Fuzzy and Uncertain Object-Oriented Database Management Systems, Proceedings of the 16th Annual Conference of North American Fuzzy Information Society , 1997, 68-72.
4. Dubois D. and Prade H., Processing Fuzzy Temporal Knowledge, IEEE Transactions on Systems, Man and Cybernetics, Vol. 19, No. 4, July/August 1989, 729-744.
5. Dubois D. and Prade H., Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.
6. Dutta S., An Event Based Fuzzy Temporal Logic, Proceedings of the 18th IEEE International Symposium on Multivalued Logic, Spain, May 1988, 64-71.
7. Dutta S., Generalized Events In Temporal Databases, Proceedings of IEEE International Conference on Data Engineering, LA, CA, USA, February 1989, 118-125.
8. Dyreson C. E. and Snodgrass R. T., Valid-Time Indeterminacy, Proceedings of IEEE International Conference on Data Engineering, Vienna, Austria, April 1993, 335-343.
9. Elmasri R. and Shamkant B. N., Fundamentals of Database Systems, The Benjamin/Cummings Company Inc., 1994.
10. McKenzie E. and Snodgrass R., Evaluation of Relational Algebras Incorporating the Time Dimension in Databases, ACM Transaction on Database Systems, 1991, 501-543.
11. Motro A., Imprecision and Incompleteness in Relational Databases: Survey, Information and Software Technology, Vol. 32, No. 9, 1990, 579-588.

12. Motro A., Accommodating Imprecision in Database Systems: Issues and Solutions, SIGMOD Record, Vol. 19, No. 4, December 1990, 69-74.
13. Petry F. E., Fuzzy Databases: Principles and Applications, Kluwer Academic Publishers, 1996
14. Snodgrass R. and Ahn I., A taxonomy of time in databases, Proceedings of the ACM-SIGMOD International Conference on Management of Data, Austin, Texas, USA, May 1985, 236-246.
15. Soo M. D., Bibliography on Temporal Databases, ACM SIGMOD Record, Vol. 20, No. 1, March 1991, 14-23.
16. Tansel A. U., Clifford J., Gadia S., Jajodia S., Segev A. and Snodgrass S., Temporal Databases: Theory, Design and Implementation, The Benjamin/Cummings Company, 1993.
17. Van Der Cruyssen B., De Caluwe R., De Tre G. and Vandycke D., A Theoretical Model for Crisp and Fuzzy Time, Proceedings of the 16th Annual Conference of North American Fuzzy Information Society , 1997, 63-67.
18. Zadeh L. A., The concept of a linguistic variable and its application to approximate reasoning, Information Science, Vol. 8., 1975, 199-249.

OUTPUT

1. Kurutach W., Integration of Fuzziness and Time in Database, Proceedings of the 20th Electrical Engineering Conference, Bangkok, November 13-14, 1997.
2. Kurutach W., Processing of Binary Temporal Constructors for Fuzzily-Bounded Time Intervals in Temporal Databases, Proceedings of the 17th Annual Conference of North American Fuzzy Information Society, Florida, USA, August 20-21, 1998.
3. Kurutach W., Handling Fuzziness in Temporal Databases, Proceedings of the 1998 IEEE International Conference on System, Man, and Cybernetics, San Diego, CA, USA, October 11-14, 1998.
4. Kurutach W., Generalized Bitemporal Constructors for A fuzzily-Bounded Time Interval Model, Proceedings of the 5th International Conference on Soft Computing, Iizuka, Fukuoka, Japan, October 16-20, 1998.
5. Kurutach W., On Incomplete Data in A Fuzzy Temporal Database Model, Proceedings of the 7th European Congress on Intelligent Techniques and Soft Computing, Aachen, Germany, September 13-16, 1999.

ពាក្យបញ្ជី

Integration of Fuzziness and Time in Databases

Werasak Kurutach

Department of Computer Engineering, Mahanakorn University of Technology

51 Cheum-Sampan Rd., Nong Chok, Bangkok 10530, THAILAND

Phone (66-2) 9883655 ext 211, Fax (66-2) 9883687, E-mail: werasak@mut.ac.th

Abstract

This paper presents a novel perspective of incorporating *fuzziness* and *time* in databases. These two features are advanced aspects of database applications and have been researched in parallel for more than a decade. A major aim of this work is to provide a (mathematical) conceptual framework, based on fuzzy set theory and the relational data model, that allows the notions of both fuzziness and time to be treated in the same model and also provide researchers in both fields of fuzzy databases and temporal databases to be aware of an important problem that can arise from the integration. In addition, a simple solution is proposed for such a problem by using a logical structure based on the relational data model to represent metadata when both fuzziness and time need to be handled.

1. Introduction

Fuzzy data are data whose meanings are imprecise, and temporal data are time values that can be stamped to some other types of data by some methods (mainly, *tuple-timestamping* and *attribute-timestamping* methods [6]). Fuzziness and time are claimed to be two important aspects of data to be represented in databases [2,10]. However, the researchers in the fuzzy database community and in the temporal database community have extended traditional data models, especially the relational data model, to treat these two features separately. Until recently, few researchers have been integrating these two aspects in a homogeneous framework [1,4]. Unfortunately, their proposed frameworks

focused only on structures and some operations of a temporal module in stead of the data model, and, thus, they did not handle an important characteristic which always arises from the integration. Such a characteristic is the time-varying meanings of fuzzy data represented in a temporal database. This property is primarily a result of the fact that the meanings of fuzzy data are context-dependent and the context always change from time to time. Hence, the meanings of fuzzy data are time-varying. The lack of the capability to model this notion can lead to serious problems - the misinterpretation of fuzzy data and the misinterpretation of query specifications.

In this paper, we propose a mathematical data model to represent fuzzy data in a temporal database, and also introduce a data structure based on the relational data model to handle the time-varying meaning of fuzzy data. This work restricts the scope that temporal data themselves are not fuzzy in order to pinpoint the problem, and we also leave other parts of the model as future work such as operations and data integrities. In section 2, we will present the mathematical-oriented conceptual model and discuss the problem by using simple examples. Section 3 proposes a data structure to handle metadata concerning the time-variant property of fuzzy data. Finally, section 4 discusses some further research in this direction and concludes this present work.

2. Mathematical Structure of Data Model with Fuzziness and Time

In our model, based on the concept of linguistic variables [11], we introduce the notion of an extension of the traditional *primitive relation scheme* which is the basic construct of a *relation scheme* [8]. This extended structure can be used as a basis to allow databases to be able to represent both quantitative and qualitative data in a unified framework. Quantitative data are data values that can be measured and expressed directly in a numerical manner (e.g. age), but can be expressed by linguistic words if fuzzy [3]. Qualitative data are ones that can't be achieved by such an approach (e.g. color), and can also be imprecise [5]. Most applications of database systems need to handle these two kinds of data. Therefore, our approach will emphasize on these two types only.

Definition 2.1. A primitive relation scheme is a six-tuple

$$(R, T, U, G, M, dom)$$

where R is a finite set of attribute names, T is a family of linguistic term sets, U is a set of basic data domains, G is a syntactic rule and M is a set of semantic rules. A function dom associates a basic data domain and its corresponding term set with each attribute name, i.e.

$$dom(A) = U_A \cup T_A$$

where $A \in R$ is an attribute name, U_A is a basic domain of A and $T_A \in T$ is a term set corresponding to U_A .

The component M is an important concept that makes the construct of the primitive relation scheme in Definition 2.1 distinct from the conventional definition. Since it is composed of functions that associate data values which may be fuzzy to their meanings which can be expressed by fuzzy sets, it allows fuzzy data to be represented in a relation. The semantic rules

in M can be classified into two main types due to the different expression of the semantics of quantitative and qualitative data. More precisely, assume that M_A be a semantic rule in M and A be an attribute name with a quantitative data domain, then

$$M_A: dom(A) \rightarrow [0,1]^{U_A}$$

In the other words, the meaning of a (fuzzy) quantitative data value can be represented by a fuzzy set which may be interpreted as a possibility distribution over its basic data domain [12].

Example. Let "Age" be an attribute name, $U_{Age} = \{1, 2, \dots, 120\}$ be its basic data domain, and $T_{Age} = \{\text{young, middle-age, old}\}$ be its term set. Then, the domain of attribute "Age" is $dom(Age) = U_{Age} \cup T_{Age} = \{1, 2, \dots, 120, \text{young, middle-age, old}\}$

The semantic rule of "Age" is defined as

$$M_{Age}: dom(Age) \rightarrow [0,1]^{(1,2, \dots, 120)}$$

The term "middle-age" can be interpreted as, for example,

$$M_{Age}(\text{middle-age}) = f(x:25,35,50,60)$$

where $(x:25,35,50,60) =$

$$\begin{cases} 1 & \text{if } 35 \leq x \leq 50 \\ 0 & \text{if } x < 25 \text{ or } x > 60 \\ (x-25)/10 & \text{if } 25 < x < 35 \\ (60-x)/10 & \text{if } 50 < x < 60 \end{cases}$$

That is, $M_{Age}(\text{middle-age}) = \{0.1/26, 0.2/27, 0.3/28, \dots, 0.2/58, 0.1/59\}$. If there is an assertion "John is a *middle-age* man", then this can be expressed in terms of possibility distribution as

Age(John)	Degree of Possibility
26	0.1
27	0.2
28	0.3
...	...
...	...
58	0.2
59	0.1

Based on Definition 2.1., if an attribute name A is relevant to qualitative data, T_A is an empty set. In the other words, $\text{dom}(A) = U_A$ is the domain of A . This is primarily due to the nature of inherent fuzziness of qualitative data that always exists at the atomic value and can't be expressed directly in a numerical way. Therefore, it requires a slightly different methodology to handle fuzzy qualitative data from fuzzy quantitative data. That is to say, if the domain of an attribute A is a set of qualitative data, then the semantic is a function

$$M_A : \text{dom}(A) \rightarrow [0,1]^{U_A}$$

and the degree of membership (or the possibility degree) is defined by the concept of a weak similarity relation

$$\text{Sim} : \text{dom}(A) \times \text{dom}(A) \rightarrow [0,1]$$

Thus, the atomic value of an object's attribute is interpreted in terms of a fuzzy set of its similar values in the same domain. This interpretation indicates to what extent other values can be considered as being able to replace the specified value. The other alternative interpretation, based on the possibility theory [12], is that each element in the support of the fuzzy set may be the attribute value of the object with the possibility degree corresponding to its membership degree

The concept of a weak similarity relation is defined and employed in our approach, because a similarity relation has been argued by many

researchers that its transitivity property is not suitable in many circumstances [9,13]. However, we contend that the property can't be ignore completely. For example, without that property, one may have $x = y$ and $y = z$, but x does not necessarily equal to z . This is counterintuitive. A weak similarity relation is defined as follows:

Definition 2.2 Let U be the universe of discourse and $\text{Sim}(U,U)$ be a fuzzy binary relation characterized by a membership function $\mu_{\text{Sim}} : U \times U \rightarrow [0,1]$. $\text{Sim}(U,U)$ is called a weak similarity relation on U if and only if

1. $\mu_{\text{Sim}}(u,u) = 1$ (reflexivity)
2. $\mu_{\text{Sim}}(u,u') = \mu_{\text{Sim}}(u',u)$ (symmetry), and
3. if $\min(\mu_{\text{Sim}}(u,u'), \mu_{\text{Sim}}(u',u'')) = 1$, then $\mu_{\text{Sim}}(u,u'') = 1$ (weak transitivity) where $u, u', u'' \in U$

Example. Let "Quality" be an attribute of employees, $U_{\text{Quality}} = \{\text{excellent, very good, good, fair, poor}\}$ be its basic data domain, and $T_{\text{Quality}} = \{\}$ be its term set. Assume that M_{Quality} is characterized by a predefined weak similarity relation (see TABLE 1). Then, for example, the semantic the term "poor" can be expressed by

$$M_{\text{Quality}}(\text{poor}) = \{1/\text{poor}, 0.5/\text{fair}, 0.1/\text{good}\}$$

	Excellent	Very Good	Good	Fair	Poor
Excellent	1	0.5	0.1	0	0
Very Good	0.5	1	0.5	0.1	0
Good	0.1	0.5	1	0.5	0.1
Fair	0	0.1	0.5	1	0.5
Poor	0	0	0.1	0.5	1.0

TABLE 1 A weak similarity relation.

The primitive relation scheme discussed so far is concentrated only on fuzziness of data to be handled. The following definition will show how the notion of temporal data can be built

into its components R and U without any change of its construct.

Definition 2.3 Let $R = \{A_1, A_2, \dots, A_n\}$ be a set of attributes in the primitive relation scheme of a relation r , and A_{tb} and A_{te} be interpreted time attributes. For both A_{tb} and A_{te} , there is the corresponding domain U_t which is a set of chronons, denoted by integers, and a variable *now*, that is, $dom(A_{tb})$ and $dom(A_{te})$ are $\{0, 1, 2, \dots, n, \text{now}\}$. R_t is said to be a relation scheme of a time-varying fuzzy relation if

$$R_t = R \cup \{A_{tb}, A_{te}\}$$

and a time-varying fuzzy relation r is defined as

$$r \subseteq U_1 \times U_2 \times \dots \times U_n \times U_t \times U_t$$

where U_i is the corresponding domain of A_i , $1 \leq i \leq n$

3. Structure of Metadata

The structure of extended relational data model incorporating both time and fuzzy data as described in the last section is not adequate because of the lack of the capability to handle the meaning of a fuzzy data value and its time-varying semantic. There needs to be a construction for a metadatabase which can be used to solved that deficiency.

Generally, *metadata* refers to data describing other data such as the meaning, context, organization, and the purposes of data. Attribute names, relational names and relation schema are simple metadata found in the conventional model and their information is kept in the *catalog* of the system. Examples of more complicated and nontrivial metadata are constraints, invalidity and completeness of the stored data [7]. A primary purpose of metadata is to provide the ability to recognize the

intentions of users and to prevent users from misinterpretation of the answer of queries. The structure proposed in this section is to be use for modeling two main kinds of metadata - the definitions of the meanings of fuzzy quantitative data and the similarity relations of fuzzy qualitative data. These metadata are not static as they have to describe time-varying fuzzy data. Therefore, both metadata and the temporal aspect have to be modeled, and they will be termed *temporal metadata*. The term *meta-database* will be used as a collection of relations that represent metadata.

For quantitative data domains, a relation in the meta-database that is used to model temporal metadata is designed to have the following schema

MetaRel(Rel_Name, Att_Name, Fuzzy_Term Definition, T_s , T_e)

where Rel_Name takes its value from the set of relation names and Att_Name from the set of attribute names. The domain of attribute Fuzzy_Term will be the set of linguistic value from the quantitative data domains which are used in the databases. A value of attribute Definition is taken from a set of trapezoidal functions defined on the basic quantitative data domain. T_s and T_e are defined in the same manner as A_{tb} and A_{te} mentioned in Definition 2.3.

The next type of metadata is the weak similarity relations of basic qualitative data domains. It is assumed that a domain has only one corresponding weak similarity relation. There are two major kinds of weak similarity relations in the context of time - one for value used in the same period and the other for value used in different periods. For the former, the schema of a relation in the meta-database is defined by

Sim_Name(Ele#1, Ele#2, Degree, T_s , T_e)

where Sim_Name is a relation name corresponding to the basic domain name, Ele#1 and Ele#2 are attributes taking values from the basic domain, Degree has a value in the unit interval $[0, 1]$ representing the degree of similarity, and T_s and T_e are defined as previously. Note that due to the symmetry property of a weak similarity relation, viz $\mu_{sim}(x, y) = \mu_{sim}(y, x)$, only one of them will be represented. Finally, for the other type, the schema of a relation in the meta-database is defined by

Sim_Name(Ele#1, T_{s1} , T_{e1} , Ele#2, T_{s2} , T_{e2} , Degree)

where Sim_Name, Ele#1, Ele#2 and Degree are defined as the former case, T_{s1} and T_{e1} represent as time interval of Ele#1, and T_{s2} and T_{e2} represent as time interval of Ele#2.

4. Discussion

In this paper, an extended structure of the relational data model has been proposed in order to incorporate both fuzziness and time in a homogenous framework. In addition, the concept of catalog has been expanded to model the property of time-varying meaning of fuzzy data. Theoretically, the justification of the suitability of our model is that the fuzziness aspect of both quantitative data and qualitative data can be handled within the same notion, and time can naturally be realized in the proposed construct without any modification. This work is emphasized on the structure of the model only as it is an initial point of the research in this direction. There are some further works needed to be accomplished. For example,

- it is possible and probably more realistic to extend that time values themselves can be fuzzy both in databases and in the meta-database,
- any other kind of context such as individual personal view of the semantics of fuzzy data

may be significant in some applications and, therefore, needs to be handled in the meta-database, and

- a query language which incorporates both fuzziness and time has to be designed, as well as algorithms of operations on both fuzzy data and temporal data.

References

- [1] Dutta S., Generalized Events In Temporal Databases, Proc. Of IEEE Int. Conf. On Data Engineering, 1989, 118-125.
- [2] Frederick E. Petry, Fuzzy Databases: Principle and Applications, Kluwer Academics Publishers, 1996.
- [3] Kurutach W. and Franklin J., A Unifying Model for Handling Uncertainty, Imprecision and Vagueness of Linguistic Data in Databases, Proc. of the 3rd Int. Workshop on Rough Sets and Soft Computing, San Jose, CA, USA, November 10-12, 1994, 531-541.
- [4] Kurutach W., Modeling Fuzzy Interval-based Temporal Information: A Temporal Database Perspective, Proc. Of the Int. Joint Conf. Of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symp., Yokohama, Japan, 20-24 March 1995, 741-748.
- [5] Kurutach W., On Imperfection Aspects of Qualitative Data in Databases, Proc. Of the 6th IFSA World Congress, Vol. 2, Sao Paulo, Brazil, July 22-28, 1995, 393-396.
- [6] McKenzie L. E. and Snodgrass R. T., Evaluation of Relational Algebras Incorporating the Time Dimension Databases, ACM Trans. on Database Systems, 1991, 501-543.
- [7] Motro A., Integrity = validity + completeness, ACM Trans. on Database Systems, Vol. 14, No. 4, 1989, 480-520.